

RESEARCH

Availability and Audit of Links in Altmetric Data Providers: Link Checking of Blogs and News in Altmetric.com, Crossref Event Data and PlumX

José Luis Ortega^{1,2}

¹ Institute for Advanced Social Studies (IESA-CSIC), ES

² Joint Research Unit "Knowledge Transfer and Innovation" CSIC-University of Córdoba, ES

jortega@iesa.csic.es

The aim of this paper is to compare and analyse the availability of blogs and news links from the three most important altmetric data providers (Altmetric.com, PlumX and Crossref Event Data, CED). In addition, the study explores the distribution of events by creation year in order to observe the coverage of old and new events. Researchers extracted 51,000 links from news and blogs from those providers. Those links were analysed with a link checker (Xenu's Link Sleuth), and the statuses of those links in 2019 January were at the center of the study. The results show that 35.6% of news in Altmetric.com are not accessible and 28.9% of blog mentions in PlumX point to a broken link. These worrying percentages of broken links are due, mainly, to the employment of third parties to supply news and blog events. Altmetric.com is the service that provides a better-balanced distribution of events, while PlumX and CED group their events around the last two years. The study concludes that these aggregators need to develop a specific policy to improve the audit of these data for research evaluation processes (saving a copy of the event, employing more frequently crawls, avoiding external providers, etc.).

Keywords: Altmetrics data providers; blogs; news; links rot; web persistence

Introduction

Altmetrics data providers are becoming indispensable tools for observing the impact of research outputs in a wide range of societal environments. From the public discussion of new results in online social networks to the mention of research articles in patent applications, these services capture the footprint of the scholarly results in web spaces representative of different social spheres. Basically, these services provide counts and links that allow users to find out when, where and how many times a document is mentioned in anywhere on the Web. In this way, web links are critical elements to verify that an event has occurred, and therefore to audit the existence of the event. Even more, without this audit possibility, it could be very hard that these platforms and their metrics could be used to support research evaluations.

It is not strange that the principal platforms have endorsed the NISO recommendations (2016) about supply transparent information and the ability to be audited for external authorities that verify the reliability of those services. For this reason, Altmetric.com does not include Mendeley readers in its Attention Score because Mendeley does not permit the site to insert a direct link that allows to verify the real number of readers (Altmetric.com 2019).

However, the audit of data supplied by altmetric providers depend, to a great extent, on the type of data gathered. The number of tweets, Mendeley readers or Wikipedia citations come from only one source which makes easier to check the real event in the original source. On the contrary, information about blogs and news comes from multiple sources which implies to pre-define a list of sources to track mentions. In face of this difficulty, many of these providers employ third parties that supply data about web events. Concretely, mentions in blogs and news are provided by external services specialized in collect scholarly blogs and media sources (clipping). For example, Altmetric.com used Moreover.com to track mentions (80%) of research articles in mainstream media, whereas PlumX fed their blog mentions from ACI Scholarly Blog Index.

The past tense is intentionally used because these third parties are now disappeared or acquired by other companies, and the links to the provided events are, in some cases, no longer available. Therefore, another problem about the accuracy of altmetric data is the persistence of the web links that allows us to verify the mention.

Literature review

In spite of the growing importance of the data providers in the altmetric studies, the literature about the functionalities and working of these platforms is not very long. Adie and Roe (2013) were the first scholars to detail how Altmetric.com tracks the mentions of papers on the Web. Trueger et al. (2015) made a critical review of the Altmetric Score, although, it

was Gumpenberger et al. (2016) who expressed the strongest criticism of the Altmetric Score as impact indicator. In the case of PlumX, Champieux (2015) and Lindsay (2016) described the utilities of the service, while Wong and Vital (2017) analysed the implementation of the tool in a specific organization. However, no studies have been yet published about the functionality of CED due mainly to this product is still in beta.

Nevertheless, many other studies have analyzed the coverage of these services, describing the proportion of altmetric events in different samples. Thelwall et al. (2013) performed the earliest distribution of metrics in Altmetric.com, finding a greater proportion of papers mentioned on Twitter and Facebook. Robinson-García et al. (2014) also analyzed the coverage of this provider and they found that 87.1% of articles had at least one tweet and 64.8% one Mendeley reader. In a similar way, Bornmann (2014) explored a set of articles from Altmetric.com and he observed that 71% of articles were tweeted and a moderated proportion of documents were mentioned in Facebook (31%). However, more recently, Thelwall (2018) studied the presence of a sample of Social Sciences, Arts and Humanities publications from Scopus in Altmetric.com. He found that 35% of articles in social sciences and 15% in arts and humanities fields were tweeted. These results suggest that only a small fraction of papers from Scopus are indexed in Altmetrics and social sciences and humanities outputs are less mentioned than STEM papers. According to PlumX, it worth mentioning the work of Torres-Salinas et al. (2017) about the collection of books. Their results showed that the distribution of events for books is rather different than for articles. Ortega (2018) used PlumX data to track the life cycle of several altmetrics, observing that the most frequent ones are Mendeley readers and Tweets.

The phenomenon of broken links or “link rot” was widely studied in the early 21st century. Those studies addressed the problem of the availability of links in a very unstable web environment. Cho and Garcia-Molina (2000) were the first researchers to study this issue. They crawled almost a million of web pages and they found that 50% of the web content roughly change after 50 days. Fetterly et al. (2003), with a higher sample, detected that 12% of the pages were inaccessible in 11 weeks. Bar-Yossef et al. (2004) devised a way to detect and measure the presence of dead pages in several samples. These types of pages followed a pronounced exponential decay. Finally, Koehler (2002, 2004) monitored the persistence of almost 400 web pages for four years. His results showed that a set of web pages declined at a half-life of about two years. Other studies have analysed in detail the persistence of links in scholarly environments. For example, Wren (2004) studied the availability of links from MEDLINE articles; Klein et al. (2014) analysed the links in the references of articles indexes in arXiv, Elsevier and PubMed Central; and Gertler and Bullock (2017) checked the links included in an academic journal. All they observed a common decay process, in which links to old pages are more likelihood to be broken than the newest ones.

Objectives

The main objective of this article is to check the availability and status of the links to blogs and news sources provided by three major altmetric data providers (Altmetric.com, CED and PlumX). The purpose is to observe to what extent these links remain active and to determine way the information supplied by the providers can be externally audited. Several research questions were formulated:

- How many links are correctly working by data provider and type of event?
- How does the persistence of links evolve in the three providers?
- How do data providers cover blogs and news according to publication date?

Methods

This experiment tries to analyse the persistence of web links using a previous sample of 51,000 links to blog posts and news from three important altmetric data providers: Altmetric.com, PlumX and Crossref Event Data (Ortega 2019). To obtain the list of links, 100,000 research articles with DOI and published after 2012 were randomly extracted from Crossref (api.crossref.org). Next, these publications were searched in Altmetric.com API (api.altmetric.com/v1/doi/) to extract the Altmetric ID. This ID was then used to scrape the Altmetric web site (www.altmetric.com/details/) and to extract the links to blogs and news. In PlumX's case, DOIs were directly used to extract information from the web site (plu.mx/plum/a/?doi=). Finally, information from CED was gathered from the API (query.eventdata.crossref.org/events?filter=obj-id:). In the three cases, several SQL scripts were written to scrape the data from websites and APIs. This process was performed during the second fortnight of August 2018.

Data extracted from Altmetric.com have an important limitation. Some news reports are locked, showing only the first four events. The proportion of hidden links would roughly be 38% (Ortega 2019). As a result, this study only covers publicly available links because the main purpose of this paper is to analyse the availability of links, hidden links are then irrelevant because a common user cannot check the status of these links.

Several links were removed from the total list of links extracted from mentions to blogs and news. This decision is due to some mentions are not actually media citations but bibliographic citations. That is, some providers consider citations from research papers as citations from blogs. This problem is especially evident in PlumX, where 23.7% of the mentions are in fact bibliographic citations, mainly from Hindawi (17.9%) and OMICS Publishing Group (3.1%) (Ortega 2019). These links were removed from the study.

Both PlumX and Altmetric.com distinguish mentions from blogs and news, showing a separate page for each type of event. This allows researchers to present the results according to each type of mention. However, CED does not distinguish between blogs and news. It defines three categories *wordpressdotcom*, *web* and *newsfeeds*. In addition, category *reddit-links* includes links from Reddit that point external sources such as blogs and news. A manual inspection evidenced that the media and blogs are equally classified as *web* or as *newsfeed*, and sometimes in both categories at the same time. Due to these classifications, the distinction between blog and news is based on the matching with the other data providers. In the case of mentions that do not match, then a manual classification is done (Ortega 2019).

Once the links from blogs and news were extracted from the altmetric data providers, a free link checker, Xenu's Link Sleuth (home.snafu.de/tilman/xenulink.html) was used to test the status of the links. The checking process was performed in 2019 January. Xenu's allows to test the status of a page according the HTTP/1.1 Status Code Definitions (Fielding et al. 2014). These codes can be grouped in five categories:

- Informational (1XX): This category indicates a provisional response, consisting only of the Status-Line and optional headers. These codes just inform that the petition has been received.
- Successful (2XX): These codes indicate that the request has succeeded.
- Redirection (3XX): This class of status code indicates that further action needs to be taken by the user agent in order to fulfil the request. Usually, it is the redirection to a new location of the resource.
- Client Error (4XX): This group of status code is intended for cases in which the client seems to have erred. Except when responding to a HEAD request, the server should include an entity containing an explanation of the error situation. These errors generally emerge when non-existent resources are requested.
- Server Error (5XX): These types of responses indicate cases in which the server is aware that it has erred or is incapable of performing the request.
- Other internal problems (12XX): In addition, the software detects another problem not specified in the HTTP/1.1 Status Code Definitions. These errors are timeouts, invalid response or connection problems.

However, the current configuration and design of some web sites causes some problems when it comes to identify a link rot and the availability of a resource. Many sites do not properly inform about the status of the request. In the case of Client Error, it is usual to redirect to a page that informs about the absence of the resource (soft 404 page) (Bar-Yossef et al. 2004). Technically, this page returns a 200 code, in spite of the content is 404. In other cases, some sites do not return any information about the disappeared pages, but they redirect, a 301 code, to the main page. This situation could disguise the actual number of broken links.

To solve this problem, a bot was designed to automatically request the page title from these links and check if the it is similar to the title reported by the altmetric provider or, instead, it includes a generic title about the news medium or blog that publish the page (for example, *Washington Post: Breaking News, World, US, DC News & Analysis*). Due to this, pages with a redirection (3XX code) and a different title were categorized as 404 error. This procedure also allows us to detect soft 404 pages when they include specific messages about an error (for example, *Oops! That page can't be found* or *Page not found*). In this case, pages with a successful code (2XX) were reclassified as 404 when they have this type of title.

It is interesting to remark another problem completely external to these data providers, but it affects the status of the request. It is the code 451, Unavailable for legal reasons. This code is multiplying due to the recent EU *General Data Protection Regulation (GDPR)*, which obliges the explicit consent of the users for placing cookies in their browser. In protest against this directive, some non-EU media have blocked the access of users from EU countries to their sites. This code was analysed apart to observe the incidence of this particular phenomenon.

Results

Distribution of status

This part studies the distribution of the pages' status in blogs and news, with the aim of detecting the proportion of broken links in each data provider.

Table 1 and **Figure 1** show the distribution of statuses of the links analysed from blogs. Altmetric.com is the site that includes more links to blogs (13,445), followed by far by PlumX (3,346) and CED (1,533). In general, the most frequent status is 3XX – Redirection, being Altmetric.com the platform that has the highest proportion (62.2%) and PlumX, the lowest one (51.9%). The presence of redirections is important as it confirms the changing nature of the web design and warns about the potential risk of broken links in altmetric data providers. The next most frequent status is 2XX – Successful, with 35.1% for CED and 27.8% for Altmetric.com. However, PlumX shows a very different proportion where the number of 4XX client error (29.0%) is greater than successful pages (18.9%). This high proportion of errors is caused by ACI Scholarly Blog Index, a former provider of academic blog posts. This service has since now disappeared (continued by ProQuest) and the links from this service are not currently active. Overall, the platforms that have a lower proportion of error codes (4XX, 5XX and 12XX) are CED (4.5%) and Altmetric.com (10.1%), whereas PlumX accumulates a 29.1% of errors.

Table 1: Distribution of status codes in blog links by altmetric data provider.

Status-Code	Altmetric.com	%	CED	%	PlumX	%
2XX – Successful	3,735	27.8%	538	35.1%	632	18.9%
3XX – Redirection	8,357	62.2%	926	60.4%	1,736	51.9%
4XX – Client Error	853	6.3%	52	3.4%	971	29.0%
5XX – Server Error	113	0.8%	5	0.3%	1	0.0%
451 – Unavailable for Legal Reasons	1	0.0%	0	0.0%	4	0.1%
12XX – Other errors	386	2.9%	12	0.8%	2	0.1%
TOTAL	13,445		1,533		3,346	

**Figure 1:** Distribution of types of status in blog links.

Table 2 and **Figure 2** show the distribution of statuses of the links captured from news media. With regard to news, Altmetric.com (15,904) and PlumX (15,585) contain a similar number of links, while CED (997) only captures a quite low amount. As in the blogs' case, the most frequent status is 3XX – Redirection, being CED the provider with the largest proportion (86.2%), followed by PlumX (58%) and Altmetric.com (54.8%). These percentages are higher than blogs, and they suggest that the news media environments are less stable. These high redirections cause lower 2XX – Successful statuses, being PlumX (19.8%) the platform with the greatest percentages, followed by CED (12.3%) and Altmetric.com (11.6%). It is interesting to notice the high proportion of 4XX errors in Altmetric.com (30.5%), which is much higher than PlumX (17.7%) and CED (1.1%). In the manner of PlumX in blogs, Altmetric.com counted on an external provider, Moreover.com, to obtain data on news mentions. This service was acquired by LexisNexis (2014) and integrated into LexisNexis Newsdesk. However, LexisNexis does not yet support many of these links, causing that they cannot be resolved now. The proportion of unresolved links is 26% (Ortega 2019). This fact could explain the high proportion of broken links in this provider. Overall, the platform that depicts a lower proportion of error codes (4XX, 5XX and 12XX) still is CED (1.5%), followed by PlumX (22.2%) and Altmetric.com (35.6%).

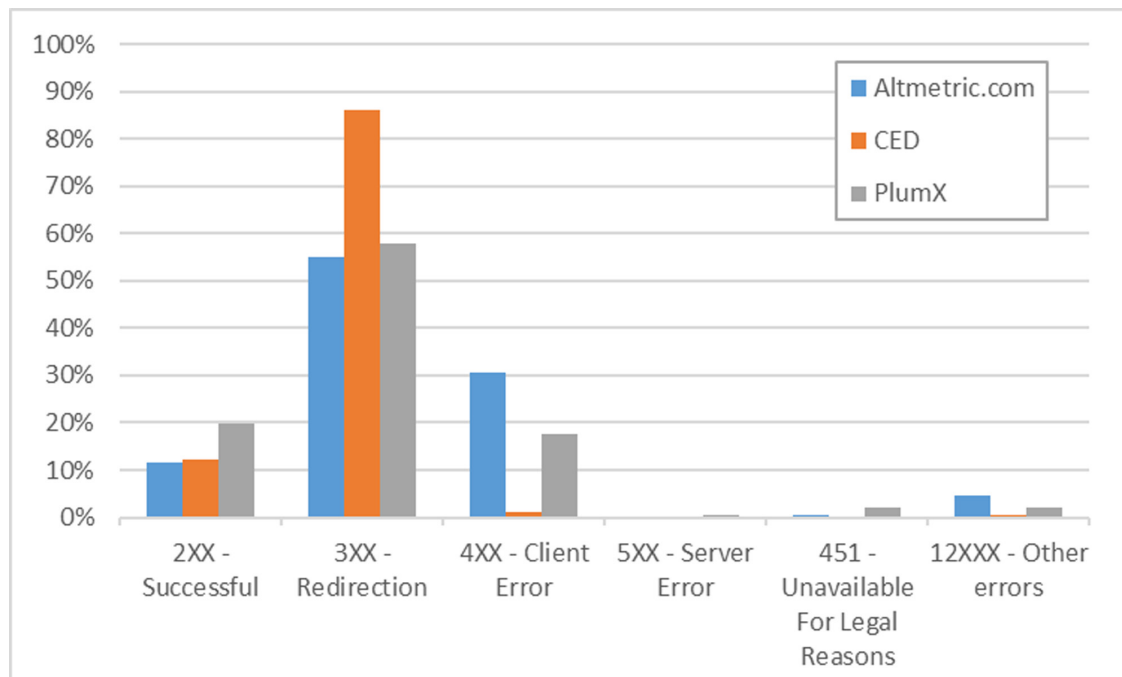
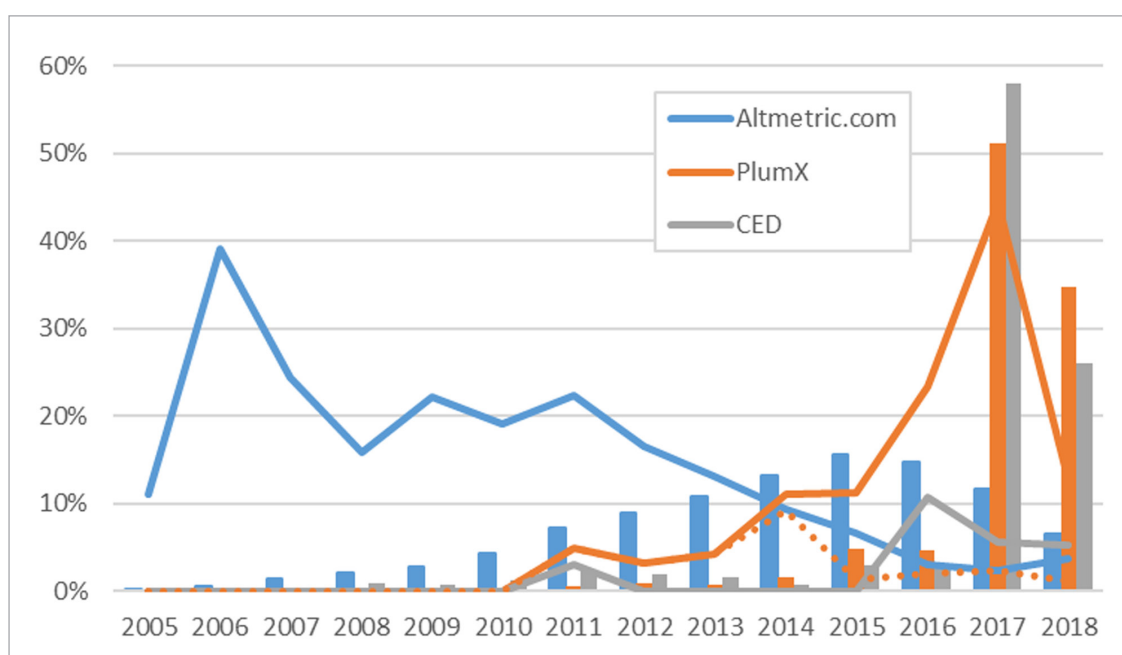
Longitudinal analysis

Another way to study the phenomenon of link rot is to observe the distribution of errors by year of the events in the three providers. The objective is to observe when the errors appear. Statuses that mark an error are 4XX – Client error, 5XX – Server error and 12XX – Other errors.

Figure 3 shows the trend of errors and events in the three data providers. Bars represent the percentage of events in each year, and the lines the percentage of errors in each year. Altmetric.com shows a better-balanced annual distribution of events, while PlumX and CED mainly gather their events around the last two years. This result confirms that the distribution of events in Altmetric.com is rather different to the other providers and that it covers events older than PlumX and CED.

Table 2: Distribution of status codes in news links by altmetric data provider.

Status-Code	Altmetric.com	%	CED	%	PlumX	%
2XX – Successful	1,803	11.6%	123	12.3%	3,091	19.8%
3XX – Redirection	8,546	54.8%	859	86.2%	9,038	58.0%
4XX – Client Error	4,751	30.5%	11	1.1%	2,763	17.7%
5XX – Server Error	26	0.2%	0	0.0%	60	0.4%
451 – Unavailable For Legal Reasons	70	0.4%	0	0.0%	322	2.1%
12XX – Other errors	708	4.5%	4	0.4%	311	2.0%
TOTAL	15,904		997		15,585	

**Figure 2:** Distribution of types of status in news links.**Figure 3:** Longitudinal distribution of the percentage of events and errors in the three providers according to blogs.

It is important to note that all of the percentage of errors distributions depict significant peaks due to one-time problems. For example, the peak in Altmetric.com is common because old blog entries could be removed or relocated in other places. In fact, the distribution follows a common decay trend in web persistence (Wren 2004; Klein et al. 2014; Gertler and Bullock 2017). However, it is more surprising that PlumX and, to a lesser extent, CED show an opposite behaviour. In the first case, it is due to ACI Scholarly Blog Index, which ceased their services for PlumX in 2018. Then, all the links provided by this service are now broken, being more than 90% of the errors in PlumX. If these links are removed, the trend is now more stable (broken line). In general, Almetric.com shows, in average, percentages of link rot in the last five years (5%) lower than PlumX (21%) but higher than CED (4%).

Another interesting fact is that PlumX and CED show no errors in the first years because they do not gather such old events. In the case of Altmetric.com, the limited number of events during this time period (2005–2009), causes that the proportion of not-found pages is higher.

Figure 4 displays the distribution of events and error statuses of links from media in the three providers. Bars represent the percentage of events in each year and the lines the percentage of errors in each year. Even though to a lesser extent than the **Figure 3**, Altmetric.com shows better coverage of old events than PlumX and CED, which mainly group their events in the two last years.

As the blogs' case, distributions show important fluctuations. Altmetric.com presents two significant peaks. The first one appears in the period 2008–2009 and it is caused by the low number of events in those moments. For example, there are two events in 2008 and one of them is broken. In 2009, there are four and three are broken links. The second peak is more important, and is due to the cease of operations of Moreover.com, a news provider. This service stopped working in 2014 and many of its links are not yet active. The broken line shows the proportion of links rot without these links. Now, the line is more stable with a low and similar proportion of broken links in each year, with percentages lower than 20%.

PlumX depicts a slight increase of errors from 2014. Many of the broken links come from US local media, which are specially covered by this service (Ortega 2019). The structure of these sites tends to be unstable and the links to their news disappear more frequently. According to CED, the number of errors is insignificant and the longitudinal distribution does not show any remarkable pattern. Overall, the percentage of broken links in the last five years confirms that Altmetric.com is the system with the highest amount of errors (31%), followed by PlumX (13%) and CED (1%).

Discussion

The persistence and availability of links is a key element in the working of altmetric data providers. They assure the truthfulness of the events captured and the existence of the source. This veracity is the element that could allow these services to be employed in research evaluation processes. This fact is even more important in the case of blogs and news because there are very different sources and the links are the only way to demonstrate that the event has occurred. On this point, the results obtained in this paper show worrying percentages of broken links in the two principal providers, 36% of media links are broken in Altmetric.com and PlumX has 29% of non-active blog links.

It is interesting that these elevated percentages are due, in both services, to the hiring of third parties to provide events. An easier way: ACI Scholarly Blog Index (PlumX) and Moreover.com (Altmetric.com) were external services that provided blog and news events to these altmetric providers. The disappearance of those platforms caused an important

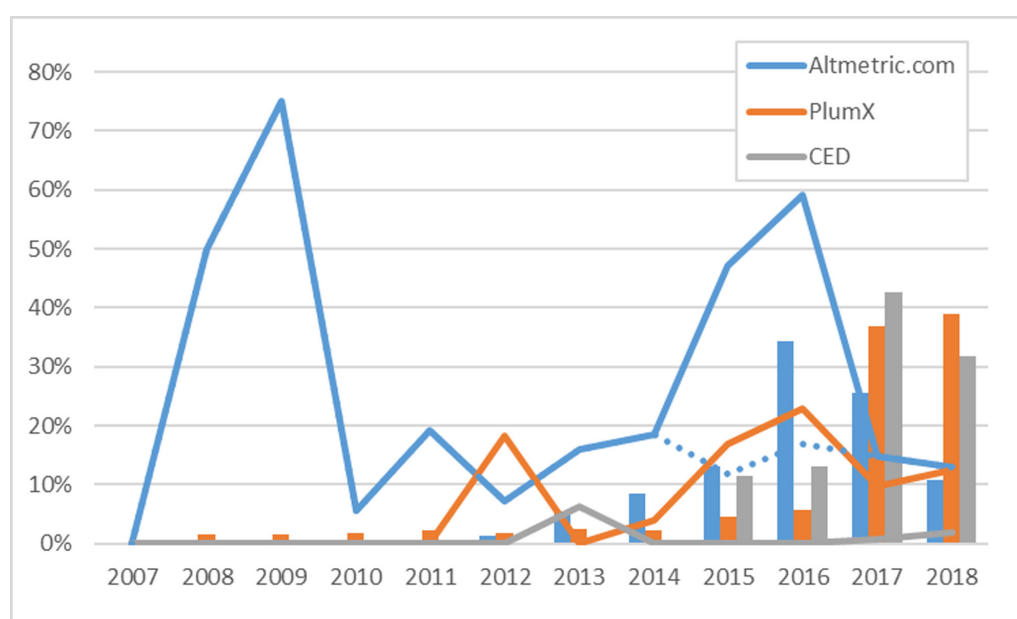


Figure 4: Longitudinal distribution of the percentage of events and errors in the three providers according to news.

percentage of their links to become inactive and therefore the information cannot be audited. This fact would serve as warning to these providers with regard to stocking up of data from unreliable or unstable sources. In contrast, CED employs a constant updating system that allows to find the most recent link version of the event (Crossref 2019). This fact causes that the levels of links rot in CED are acceptable (<5%).

The high proportion of redirections (3XX), more than 50% in the three providers, informs us about the current unstable nature of the Web and how the sites usually restructure their contents. However, these redirections always imply risk of broken links because the contents are not always correctly relocated. Other important risk is caused by the 451 status, which blocks the access by legal reasons. The results show that this type of errors is infrequent, and in the worst-case scenario, news from PlumX, the result is only 2%.

From the longitudinal view, results show that Altmetric.com is the provider that captures more events from old pages, which is also a factor that increases the likelihood of broken links. Both Altmetric.com and PlumX depict important peaks of error pages, mainly due to the disappearance of their data providers. Besides this, all the providers describe a constant trend throughout the years. The fact that these lines do not follow a usual decay process (Wren 2004; Klein et al. 2014; Gertler and Bullock 2017) could be due to the small size of the sample. Except for Altmetric.com, which seems following a descendent behaviour.

Conclusion

The obtained results lead us to conclude that the percentage of broken links is worrying in the two principal altmetric data providers, being 36% for the news of Altmetric.com and 29% for the blogs of PlumX. These elevated percentages are due to links provided by third parties. Those results lead us to recommend developing a more frequent system of gathering events and improving the way of auditing their data (saving a copy of the event, employing more frequently crawls, avoiding external providers, etc.).

Another conclusion is that the longitudinal analysis demonstrates that the distribution of broken links is not uniform, as it presents important peaks in the three providers. These peaks are mainly due to external data sources.

The results also allow us to conclude that Altmetric.com has a better coverage of old events than PlumX and CED. This fact should be considered when altmetric studies attempt to analyse old events.

Competing Interests

The author has no competing interests to declare.

References

- Adie, E., & Roe, W.** (2013). Altmetric: Enriching scholarly content with article-level discussion and metrics. *Learned Publishing*, 26(1), 11–17. DOI: <https://doi.org/10.1087/20130103>
- Altmetric.com** (2019). *How is the Altmetric Attention Score calculated?* Retrieved from <https://help.altmetric.com/support/solutions/articles/6000060969-how-is-the-altmetric-attention-score-calculated->
- Bar-Yossef, Z., Broder, A. Z., Kumar, R., & Tomkins, A.** (2004, May). Sic transit gloria telae: Towards an understanding of the web's decay. In *Proceedings of the 13th international conference on World Wide Web* (pp. 328–337). ACM. DOI: <https://doi.org/10.1145/988672.988716>
- Bornmann, L.** (2014). Validity of altmetrics data for measuring societal impact: A study using data from Altmetric and F1000Prime. *Journal of Informetrics*, 8(4), 935–950. DOI: <https://doi.org/10.1016/j.joi.2014.09.007>
- Champieux, R.** (2015). PlumX. *Journal of the Medical Library Association: JMLA*, 103(1), 63. DOI: <https://doi.org/10.3163/1536-5050.103.1.019>
- Cho, J., & Garcia-Molina, H.** (2000, Sept.). The evolution of the web and implications for an incremental crawler. In *Proc. of the 26th International Conference on Very Large Databases* (pp. 200–209). Morgan Kaufmann Publishers, Inc.
- Crossref.** (2019). Appendix: NISO Altmetrics Code of Conduct. *Event Data User Guide*. Retrieved from <https://www.eventdata.crossref.org/guide/app-niso/>
- Fetterly, D., Manasse, M., Najork, M., & Wiener, J.** (2003, May). A large-scale study of the evolution of web pages. In *Proceedings of the 12th international conference on World Wide Web* (pp. 669–678). ACM. DOI: <https://doi.org/10.1145/775152.775246>
- Fielding, R., Nottingham, M., & Reschke, J.** (2014). *Hypertext Transfer Protocol (HTTP/1.1): Caching* (No. RFC 7234). Retrieved from <https://tools.ietf.org/html/rfc7234>. DOI: <https://doi.org/10.17487/rfc7234>
- Gertler, A. L., & Bullock, J. G.** (2017). Reference rot: An emerging threat to transparency in political science. *PS: Political Science & Politics*, 50(1), 166–171. DOI: <https://doi.org/10.1017/S1049096516002353>
- Gumpenberger, C., Glänzel, W., & Gorraiz, J.** (2016). The ecstasy and the agony of the altmetric score. *Scientometrics*, 108(2), 977–982. DOI: <https://doi.org/10.1007/s11192-016-1991-5>
- Klein, M., Van de Sompel, H., Sanderson, R., Shankar, H., Balakireva, L., Zhou, K., & Tobin, R.** (2014). Scholarly context not found: One in five articles suffers from reference rot. *PloS one*, 9(12), e115253. DOI: <https://doi.org/10.1371/journal.pone.0115253>
- Koehler, W.** (2002). Web page change and persistence—A four-year longitudinal study. *Journal of the American Society for Information Science and Technology*, 53(2), 162–171. DOI: <https://doi.org/10.1002/asi.10018>

- Koehler, W.** (2004). A longitudinal study of Web pages continued: A consideration of document persistence. *Information Research*, 9(2), paper 174.
- LexisNexis.** (2014). *LexisNexis Acquires Moreover Technologies*. Retrieved from <https://www.lexisnexis.com/en-us/about-us/media/press-release.page?id=1413560584448681>
- Lindsay, J. M.** (2016). PlumX from Plum Analytics: Not Just Altmetrics. *Journal of Electronic Resources in Medical Libraries*, 13(1), 8–17. DOI: <https://doi.org/10.1080/15424065.2016.1142836>
- NISO.** (2016). *Outputs of the NISO Alternative Assessment Metrics Project. NISO RP-25-2016*. Retrieved from: https://groups.niso.org/apps/group_public/download.php/17091/NISO%20RP-25-2016%20Outputs%20of%20the%20NISO%20Alternative%20Assessment%20Project.pdf
- Ortega, J. L.** (2018). The life cycle of altmetric impact: A longitudinal study of six metrics from PlumX. *Journal of Informetrics*, 12(3), 579–589. DOI: <https://doi.org/10.1016/j.joi.2018.06.001>
- Ortega, J. L.** (2019). The coverage of blogs and news in the three major altmetric data providers. In *Proceeding of the 17th International Conference on Scientometrics & Informetrics*. <https://osf.io/pd27h/>
- Robinson-García, N., Torres-Salinas, D., Zahedi, Z., & Costas, R.** (2014). Nuevos datos, nuevas posibilidades: Revelando el interior de Altmetric.com. *El Profesional de la Información*, 23(4), 359–366. DOI: <https://doi.org/10.3145/epi.2014.jul.03>
- Thelwall, M.** (2018). Altmetric prevalence in the social sciences, arts and humanities: Where are the online discussions? *Journal of Altmetrics*, 1(1). DOI: <https://doi.org/10.29024/joa.6>
- Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R.** (2013). Do altmetrics work? Twitter and ten other social web services. *PloS one*, 8(5), e64841. DOI: <https://doi.org/10.1371/journal.pone.0064841>
- Trueger, N. S., Thoma, B., Hsu, C. H., Sullivan, D., Peters, L., & Lin, M.** (2015). The altmetric score: A new measure for article-level dissemination and impact. *Annals of Emergency Medicine*, 66(5): 549–553. DOI: <https://doi.org/10.1016/j.annemergmed.2015.04.022>
- Wong, E. Y., & Vital, S. M.** (2017). PlumX: A tool to showcase academic profile and distinction. *Digital Library Perspectives*, 33(4), 305–313. DOI: <https://doi.org/10.1108/DLP-12-2016-0047>
- Wren, J. D.** (2004). 404 not found: The stability and persistence of URLs published in MEDLINE. *Bioinformatics*, 20(5), 668–672. DOI: <https://doi.org/10.1093/bioinformatics/btg465>

How to cite this article: Ortega, J. L. (2019). Availability and Audit of Links in Altmetric Data Providers: Link Checking of Blogs and News in Altmetric.com, Crossref Event Data and PlumX. *Journal of Altmetrics*, 2(1): 4. DOI: <https://doi.org/10.29024/joa.14>

Submitted: 13 February 2019

Accepted: 02 July 2019

Published: 18 July 2019

Copyright: © 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



Journal of Altmetrics is a peer-reviewed open access journal published by Levy Library Press.

OPEN ACCESS